Run 2 Data Handling at CDF

J. Tseng
Massachusetts Institute of Technology

Fermilab Director's Review
Run 2 Computing
4 June 2002

- ▷ Introduction
- ▷ Legacy system
- ▷ Enstore migration
- ▷ dCache development
- ▷ SAM/Grid
- ▷ Conclusion

Data Handling: User Perspective

▷ Simplest use, ideal case:

```
mod talk DHInput
  input dataset myset
exit
ev begin
```

▷ Data handling system:

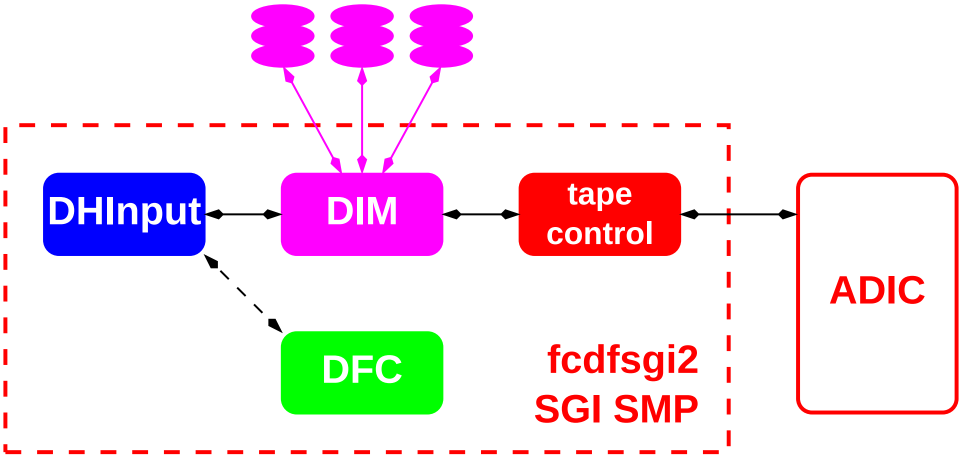
1. abstraction → physical entities
2. locates physical entities
3. delivers entities to accessible location
4. may arrange computing resources

▷ The Good News

○ already doable

○ physicists already analyzing data

Original CDF DH System



The Caveats

▷ reliability/maturity
forestalled following developments...

▷ centralization
bottlenecks, single points of failure

▷ scalability
all on a few big, central machines

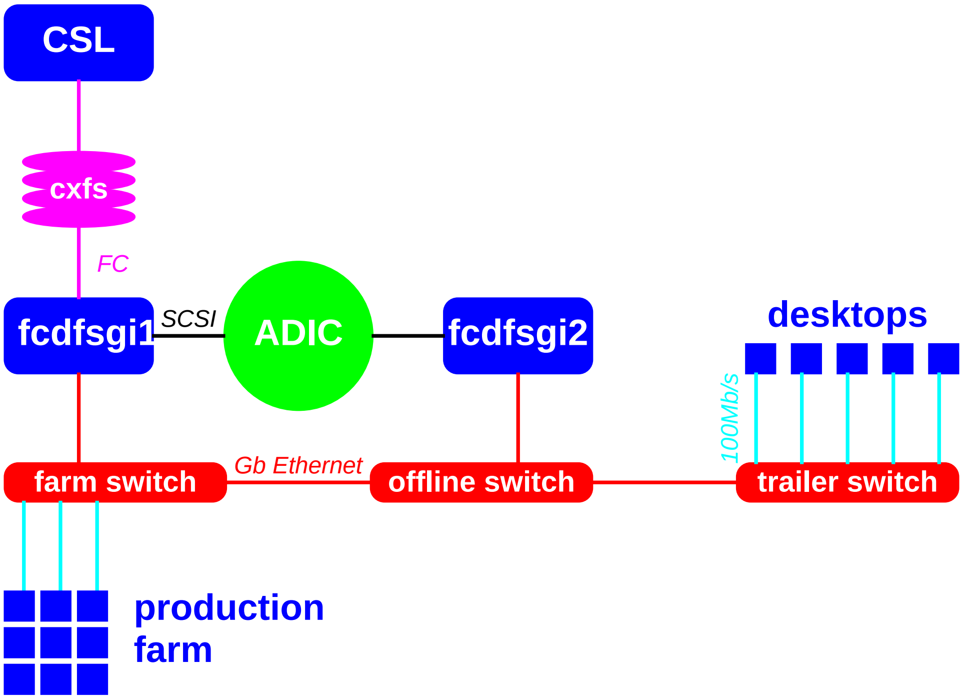
▷ adaptability
narrow technology choices

Original CDF DH System (II)

▷ Works: small scale, controlled environments

√ raw data logging (write only)

√ production farm (read/write)



▷ Tape in data path ⇒ large access latencies

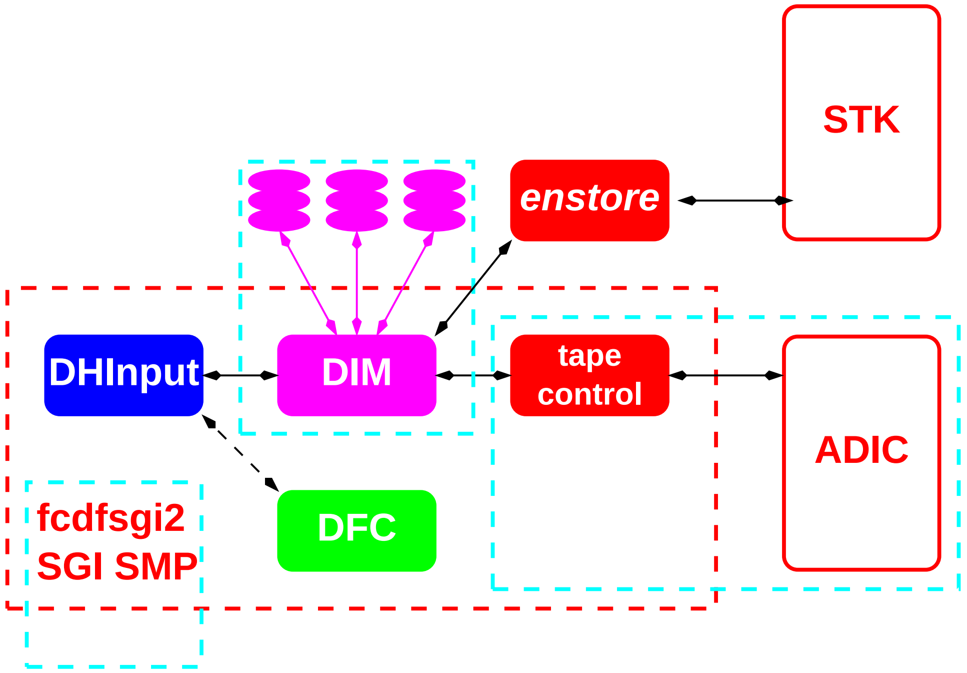
▷ `fcdfsgi1` load

▷ Problematic operations in production

⇒ Data Handling Review: Sep - Dec 2001

New CDF DH System

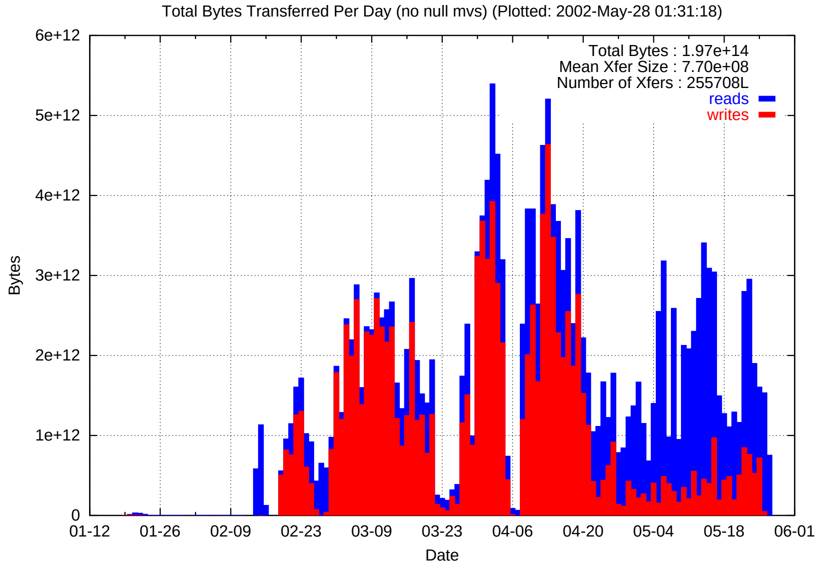
1. tape handling software debugging
   (M. Votava, M. Mengel, L. Piccoli)
2. Enstore: network-attached tapes
   Functional divisions have aided migration



3. dCache: network-attached disk cache
4. distributed computing power
   - ▷ rootd for static files on fileservers
   - ▷ infrastructure reliability even more essential as system grows more complex
   - ▷ SAM/Grid

Enstore Migration

▷ CDFEN: STK silo, 10 mover nodes (18 Feb)


Total Bytes Transferred Per Day (no null mvs) (Plotted: 2002-May-28 01:31:18)

▷ all data in ADIC copied to CDFEN

- √ avg write 14 MB/s (65 MB/s peak)
- √ total 100 TB by 22 April (3 months)

▷ continued writing:

- √ production farms (24 April)
- √ raw data logging (1 May)

▷ Enstore error handling improved (`fcdfsgi2` freezes)

▷ great collaboration with ISD

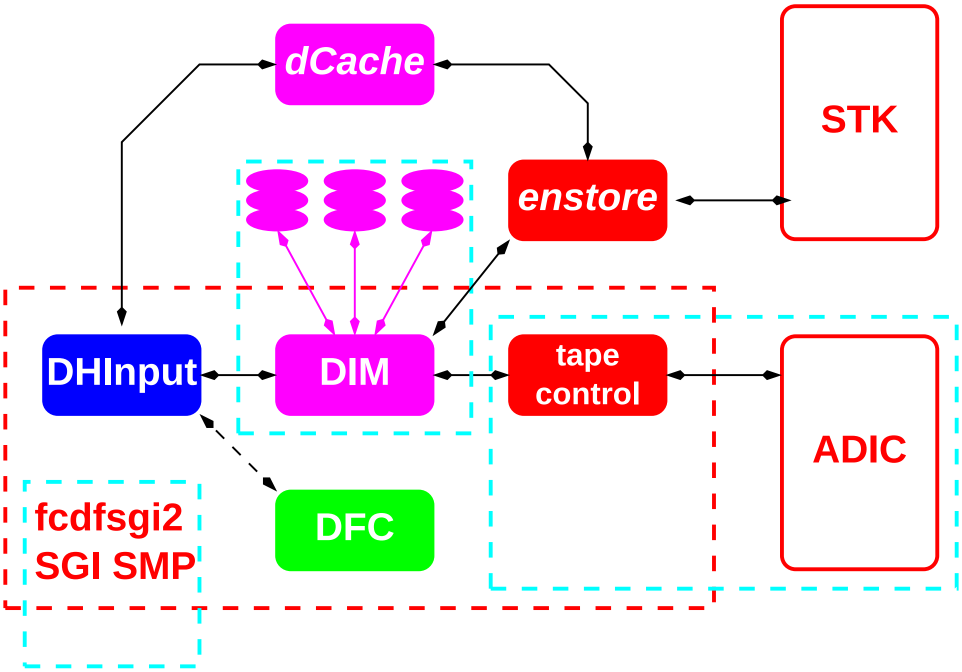
⇒ much smoother operation

DIM Replacement

▷ DIM review (CDF 5917)

- SMP-centric
- fragile, high maintenance
- CDF-only support

⇒ dCache: network-attached disk caches



▷ adaptable to distributed computing model

▷ commodity disk drives
(aim to support several pool varieties)

▷ `TDCacheFile`: simple adaptation to CDF

▷ write pools take tape out of data path

⇒ lower latency from generation to user

DIM Replacement (II)

▷ working closely with ISD

▷ evaluation cluster

- √ reads via file copy (`kftp`, `dccp`)
- √ direct random access (`libdcap`)
- √ writes via file copy
- √ cache behavior

▷ evaluate TBFS from CAF cluster

- √ behavior with larger pool (∼2 TB)
- √ selectable cache mechanism in `DHInput`
- √ load balancing

▷ prototype system with TBFS from CAF

- 4 TBFS for Stage 1 CAF (8 TB)
- Gigabit Ethernet connections
- adding DH, "power" users

▷ in use for physics

dCache in CDF

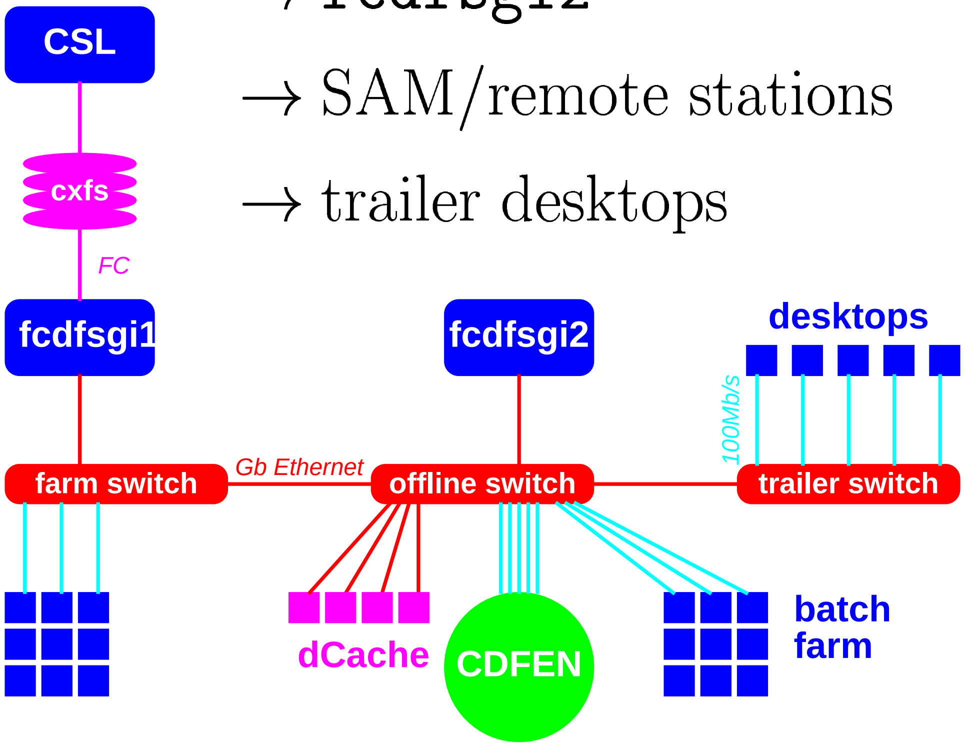
▷ Read pools first in production

→ CAF farm

→ `fcdfsgi2`

→ SAM/remote stations

→ trailer desktops



▷ Small write pools for tests

- ○ write pools demand greater robustness
- ○ main issue: authentication

▷ further wish list:
file replication, load limitation

dCache: starting beta tests
⇒ tune up for production use

SAM/Grid

▷ Data distribution to remote sites:
review by CDF UK institutions

- DFC metadata → SAM db
- SAM use of dCache
- archive read/write
- consistent metadata
- ○ prototype: DH, “power” users

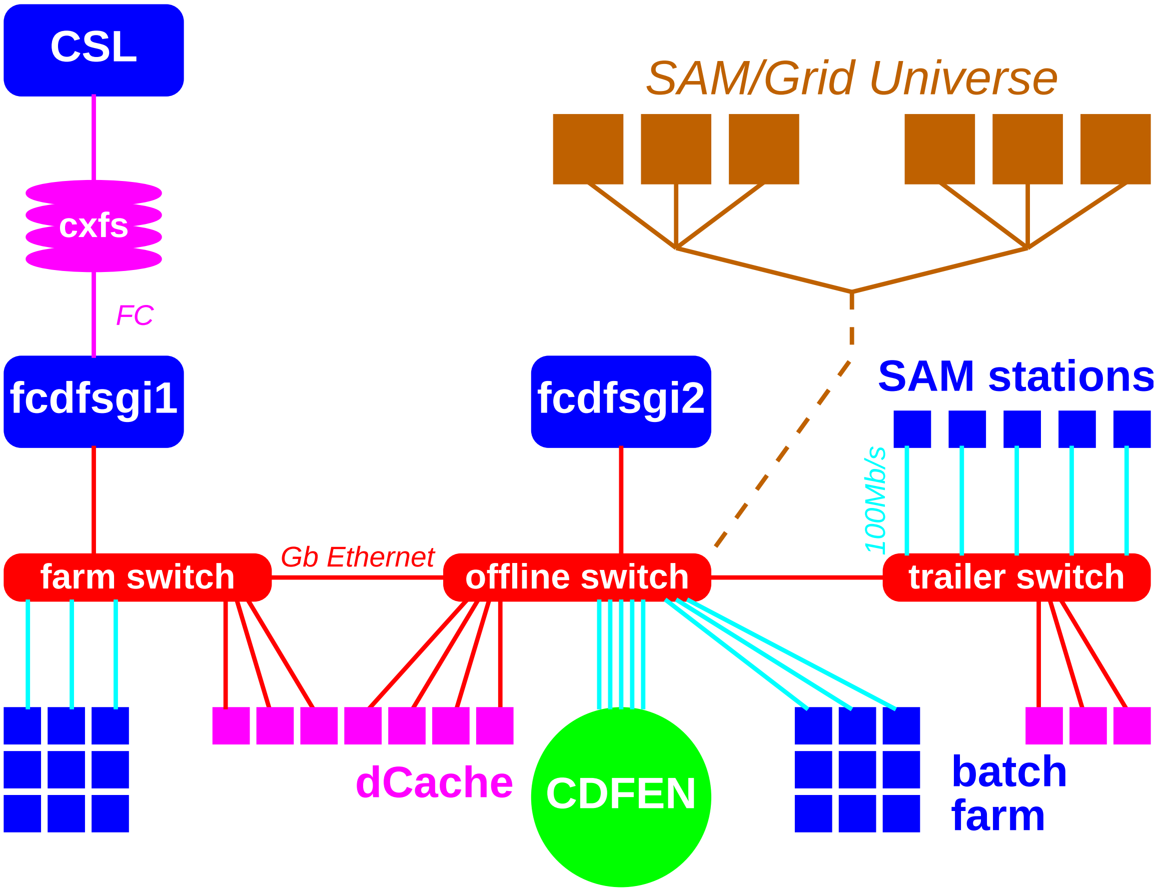
▷ Use on-site: DIM review (CDF 5917)
depended on UK review

⇒ long-term direction for CDF data handling

▷ Rick St Denis will discuss in more detail

Future DH/Grid Direction

▷ distributed peers rather than hierarchy



- current focus: singleton stations

○ design deployment: clusters/batch farms

○ WAN impact?

Future DH/Grid Direction (II)

▷ A cluster is a large error amplifier

⇒ Reliability of infrastructure essential to usability of a complex system, *e.g.*, Grid

- ∘ Enstore for network-attached tape access
- ∘ dCache for network-attached disk caching

▷ Encapsulation eases transition to using new infrastructure technologies, rather than fixing technology choices for years at a time

▷ "SAMified" Grid or "Gridified" SAM?

▷ Grid is a toolkit for data access

- ∘ take inspiration from the web: many ways to access shareable hypertext (Netscape, **`wget`**, **`lynx`**, local access)
- ∘ different means for different purposes, capabilities

▷ The only way to ensure that Grid is useful for anyone is to make sure it is useful for someone

Conclusion

- Abstractions in Run 2 DH are already an improvement over Run 1 file lists
- CD expertise has been and will be central to increasing success of CDF data handling
- reliability has increased as modular components incorporated
- SAM/Grid is the long-term direction for CDF data handling
- active Grid participation: we benefit from even wider computing expertise, and Grid gets a real-world thrashing